

02 228
YOR

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日

Date of Application:

2003年 1月30日

出願番号

Application Number:

特願2003-022380

[ST.10/C]:

[JP2003-022380]

出願人

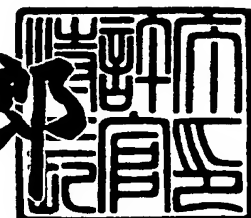
Applicant(s):

インターナショナル・ビジネス・マシーンズ・コーポレーション

2003年 5月20日

特許庁長官
Commissioner,
Japan Patent Office

太田信一郎



出証番号 出証特2003-3037026

【書類名】 特許願

【整理番号】 JP902228

【あて先】 特許庁長官 殿

【国際特許分類】 G06F 17/30

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

 【氏名】 安部 麻里

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

 【氏名】 堀 雅洋

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

 【氏名】 小野 康一

【発明者】

 【住所又は居所】 神奈川県大和市下鶴間 1 6 2 3 番地 1 4 日本アイ・ピー・エム株式会社 東京基礎研究所内

 【氏名】 小柳 光生

【特許出願人】

 【識別番号】 390009531

 【氏名又は名称】 インターナショナル・ビジネス・マシーンズ・コーポレーション

【代理人】

 【識別番号】 100086243

 【弁理士】

 【氏名又は名称】 坂口 博

 【連絡先】 0 4 6 - 2 1 5 - 3 3 1 8、3 3 2 5、3 4 5 5

【選任した代理人】

【識別番号】 100091568

【弁理士】

【氏名又は名称】 市位 嘉宏

【選任した代理人】

【識別番号】 100108501

【弁理士】

【氏名又は名称】 上野 剛史

【手数料の表示】

【予納台帳番号】 024154

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9706050

【包括委任状番号】 9704733

【包括委任状番号】 0207860

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 構造パターン候補を生成する方法、システムおよびプログラム

【特許請求の範囲】

【請求項 1】

構造化文書内の要素または要素群を指す構造パターンの候補を生成する方法であって、

構造化文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を用意するステップと、

ユーザにより指示された、前記構造化文書内の要素または要素群を指す構造パターンを受付けるステップと、

前記構造パターンを構成する項のうち編集対象となる編集対象項を決定するステップと、

前記文書論理構造情報を基に、前記編集対象項を異なる表現の項に編集することにより構造パターン候補を生成するステップと、

前記表示指標にしたがって、生成された前記構造パターン候補を並び替え、構造パターン候補一覧を生成するステップと、

を含む構造パターン候補を生成する方法。

【請求項 2】

前記編集対象項を決定するステップは、前記構造パターンについてユーザにより指示された編集希望項を受付けるステップと、当該編集希望項を編集対象項として決定するステップとを含む、請求項 1 記載の構造パターン候補を生成する方法。

【請求項 3】

前記表示指標は、前記構造化文書において、前記生成された構造パターン候補が指す要素群とユーザにより指示された前記構造パターンが指す要素群との包含関係を示す指標を含む、請求項 1 または 2 いずれかの請求項記載の構造パターン候補を生成する方法。

【請求項 4】

前記表示指標は、生成された前記構造パターン候補の特徴を示すカウント可能

な特徴量を含む、請求項 1 乃至 3 いずれかの請求項記載の構造パターン候補を生成する方法。

【請求項 5】前記文書論理構造情報を用意するステップは、前記構造化文書および当該構造化文書の文書型定義の少なくとも一方を用意するステップを含む、請求項 1 乃至 4 いずれかの請求項記載の構造パターン候補を生成する方法。

【請求項 6】

構造化文書内の要素または要素群を指す構造パターンの候補を生成するシステムであって、

構造化文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を記憶する記憶装置と、

ユーザにより指示された、前記構造化文書内の要素または要素群を指す構造パターンを受付ける手段と、

前記構造パターンを構成する項のうち編集対象となる編集対象項を決定する手段と、

前記文書論理構造情報を基に、前記編集対象項を異なる表現の項に編集することにより構造パターン候補を生成する手段と、

前記表示指標にしたがって、生成された前記構造パターン候補を並べ替え、構造パターン候補一覧を生成する手段と、

を含む構造パターン候補を生成するシステム。

【請求項 7】

前記編集対象項を決定する手段は、前記構造パターンについてユーザにより指示された編集希望項を受付ける手段と、当該編集希望項を編集対象項として決定する手段とを含む、請求項 6 記載の構造パターン候補を生成するシステム。

【請求項 8】

前記表示指標は、前記構造化文書において、前記生成された構造パターン候補が指す要素群とユーザにより指示された前記構造パターンが指す要素群との包含関係を示す指標を含む、請求項 6 または 7 いずれかの請求項記載の構造パターン候補を生成するシステム。

【請求項 9】

前記表示指標は、生成された前記構造パターン候補の特徴を示すカウント可能な特徴量を含む、請求項 6 乃至 8 いずれかの請求項記載の構造パターン候補を生成するシステム。

【請求項 1 0】前記記憶装置は、前記文書論理構造情報として、前記構造化文書および当該構造化文書の文書型定義の少なくとも一方を記憶する、請求項 6 乃至 9 いずれかの請求項記載の構造パターン候補を生成するシステム。

【請求項 1 1】

構造化文書内の要素または要素群を指す構造パターンの候補を生成するためのプログラムであって、該プログラムがコンピュータに、

構造化文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を用意する機能、

ユーザにより指示された、前記構造化文書内の要素または要素群を指す構造パターンを受付ける機能、

前記構造パターンを構成する項のうち編集対象となる編集対象項を決定する機能、

前記文書論理構造情報を基に、前記編集対象項を異なる表現の項に編集することにより構造パターン候補を生成する機能、および

前記表示指標にしたがって、生成された前記構造パターン候補を並べ替え、構造パターン候補一覧を生成する機能を実現させる構造パターン候補生成プログラム。

【請求項 1 2】

XML 文書内の要素または要素群を指す X P a t h の候補を生成する方法であって、

XML 文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を用意するステップと、

ユーザにより指示された、前記 XML 文書内の要素または要素群を指す X P a t h を受付けるステップと、

前記 X P a t h を構成するロケーションステップのうち編集対象となるロケーションステップを決定するステップと、

前記文書論理構造情報を基に、決定された前記ロケーションステップを異なる表現のロケーションステップに編集することによりX P a t h 候補を生成するステップと、

前記表示指標にしたがって、生成された前記X P a t h 候補を並び替え、X P a t h 候補一覧を生成するステップと、

を含むX P a t h 候補を生成する方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、構造化文書内の要素または要素群を指す構造パターンの候補を生成する方法、システムおよびプログラムに関する。

【0002】

【従来の技術】構造化文書とは、文書要素と要素間の論理的な関係（文書論理構造）を伴う文書である。構造化文書としては、例えばS G M L（Standard Generalized Markup Language）文書あるいはX M L（eXtensible Markup Language）文書がある。S G M LはI S O（International Standard Organization）で定められた規格である。一方、X M Lは、W 3 C（World Wide Web Consortium）で定められた規格である。X M Lは、S G M Lの一部の機能を継承しつつインターネットにおける文書フォーマットの実質的な標準であるH T M L（Hyper Text Markup Language）による運用の問題を踏まえて制定されものである。

【0003】ところで、このような構造化文書では、文書の変換・加工処理が重要である。例えば、近年、P D A（Personal Digital Assistants）や携帯電話にもウェブ・ブラウザが搭載されるようになったが、これら小型携帯端末には、一般に、表示領域の限られた画面や低速な通信手段しか装備されない。そのため、X M L準拠のアノテーション（注釈）言語を使用して、デスクトップP C用H T M L文書から小型ディスプレイでの表示に適した内容だけを抽出する加工処理が行われる。また、X M Lで記述されたコンテンツをH T M L対応のブラウザに表示し、あるいは企業間で異なるフォーマットのデータを容易に交換することが強く望まれている。そのため、X S L T（XSL Transformations）を用いて、X M Lで記述されたコンテンツをH T M LやP D Fによる表示形式へ変換する処理や

、XML文書の変形が行われる。なお、XML文書用のスタイルシート言語であるXSL (eXtensible Stylesheet Language) は、書式化の対象であるXML文書の構造変換と、得られる書式の意味を記述するための語彙から構成されているが、上述したXSLTは、このうち構造変換の部分を実現する技術を指す。

【0004】このような構造化文書の変換・加工処理において、変換・加工処理の対象となる要素は、構造パターンにより指示される。ここで、構造パターンとは、構造化文書の文書論理構造中の要素を指し示す表現を意味するが、本明細書において用いられる用語「構造パターン」とは、対象とする構造の階層／階層群を指定する階層指定子と、当該階層において選択されるべき要素／要素群を指示する要素パターンとからなる階層指示項の、列からなる表現である。なお、これ以後は、階層指示項を単に「項」と呼ぶ。例えば、XMLの構造パターンとしてW3C (World Wide Web Consortium) により定められたXPath (XML Path Language)が挙げられる。XPathでは、項は、ロケーションステップと呼ばれる。なお、XPathについては、構造パターンの具体例として、後に詳述する。

【0005】しかしながら、構造パターンによる処理対象の指示には次の問題がある。すなわち、対象となる構造化文書が変更されると、構造パターンは元の要素を指さなくなることが起こり得るという問題である。そのため、元の構造化文書の変更にあわせて、構造パターンを変更する必要がある。しかしながら、かかるメンテナンスは人手によって行われるため、多大な労力を要する。対象となる構造化文書が、インターネットを介してアクセスされるHTML文書等の場合、その内容は日々変更され変化していくため、この問題による影響は大きい。

【0006】かかる問題を回避する従来技術として、指示対象となる各要素に一意的識別子を付与する方法がある。例えば、W3C (World Wide Web Consortium) によって開発されたWebオーサリングツールであるAmayaでは、ID属性を用いて要素を指定する（非特許文献1参照）。このID属性による要素指定は、その要素自身が削除されない限り、文書の変更等の影響を受けない。しかしながら、IDを利用した要素指定は編集コストがかかり、現実的な解決方法とはいえない。

【0007】また、別の従来技術として、文書変更後に構造パターンを当該変更に対応させる方法がある（非特許文献2参照）。この方法では、元の要素は、当該要素の名前を手掛かりとして、文書の階層構造を所定の方針に従い辿ることにより検索される。しかしながら、所定の方針、すなわち、検索範囲や検索順序を指示する所定の方針を、どのように決定するかについて問題が残る。

【0008】ところで、構造化文書内の特定の要素を指す構造パターンは1つとは限らない。具体例を挙げて以下に説明する。図15（a）は、ある構造化文書の階層構造をツリー形式で表したものである。ドキュメント順、すなわち文書中の出現順位は、要素R130、要素A131、要素B132、要素C133、要素D134である。図15（a）中に示される要素D134を指し示す構造パターンは、1つには、親から子へと順次辿っていくものが考えられる。すなわち、「Rという名前の要素の、Aという名前の子要素の、更に、Bという名前の子要素の、Dという名前の子要素」と表現される構造パターンである。後に詳述するXPathの表記法に従えば、「/child::R[1]/child::A[1]/child::B[1]/child::D[1]」と表せる（第1の構造パターン）。また、要素D134を指し示す他の構造パターンとして、要素R130から一足飛びに要素D134を指示することもできる。この場合、要素R130から見た要素D134はその子孫にあたるため、「Rという名前の要素の、Dという名前の孫要素」と表現できる。XPathの表記法に従えば「/child::R[1]/descendant::D[1]」と表せる（第2の構造パターン）。同様に、要素B132のみを飛び越した場合は、「Rという名前の要素の、Aという名前の子要素の、Dという名前の孫要素」と表現できる。XPathの表記法に従えば「/child::R[1]/child::A[1]/descendant::D[1]」と表せる（第3の構造パターン）。

【0009】上記3つの構造パターンは、何れも要素D134を指し示すが、後者2つの構造パターンは、上述した文書の変更に対して耐久性を有する。例えば、文書の変更により要素B132が削除され、その結果、要素C133、要素D134が要素A131の子要素になったとする（図15（b）参照）。この場合、第1の構造パターンは、文書内に存在しない要素を指すこととなる。一方、後者2つの構造パターンは、依然として要素D134を指し続ける。更に、後者2

つの構造パターンについても、その耐久性の内容に違いがある。例えば、要素D 1 3 4と同じ名前の要素D 1 3 5を、要素R 1 3 0の真下に出現順が要素A 1 3 1よりも先となるように追加したとする（図15（c）参照）。この場合、第2の構造パターンは、要素R 1 3 0の、Dという名前の最初の孫要素を指すため、新たに追加された要素D 1 3 5を指してしまう。これに対し、第3の構造パターンは、要素R 1 3 0の、要素A 1 3 1の、Dという名前の最初の孫要素を指すため、元の要素D 1 3 4を正確に指し続ける。

【0010】このように、耐久性ある構造パターンを用いることにより、上述の文書の変更の問題を解決することができる。しかしながら、耐久性ある構造パターンは、親から子へと順次辿っていく構造パターン（以下「固定パス」と呼ぶ）のように単純なものではなく、その作成は困難を要する。また、耐久性ある構造パターンは幾通りも存在するため、将来起こり得る文書の変更に最も適した構造パターンを選択することは難しい。

【0011】それにもかかわらず、従来技術では、耐久性のある構造パターンを作成するための編集環境は提供されていない。例えば、X S L T編集システムとして、e X c e l o n社の「eXcelon Stylus」（<http://www.exceloncorp.com>）、A l t o v a社の「XML Spy」（<http://www.xmlspy.com>）、I B M社の「IBM X S L Editor」（<http://www.alphaworks.ibm.com/tech/xsleditor>）等がある。これら編集環境では、X P a t hを自動生成する機能はあるものの、生成されるX P a t hは、親から子へと順次辿っていく単純な固定パスに限定されている。したがって、耐久性ある構造パターンを生成するためには、ユーザは、直接文字列を入力し、あるいは、メニューによる補助的なツールを利用して、固定パスを編集する必要がある。そのため、幾通りもある複雑な構造パターンを生成することは難しい。また、ユーザには構造パターンに関する詳しい知識が要求される。

【0012】また、ユーザに検索結果の1例を例示させ、ユーザの例示によって得られた構造化文書の部分構造が検索結果に含まれるような構造パターンを自動作成する技術がある（特許文献1参照）。しかしながら、当該技術は、例示によって得られた構造化文書の部分構造を単に含むか含まないかで構造パターンが正しい、正しくないとしており、耐久性のある構造パターンの作成を積極的に

支援するものではない。これは、当該技術が構造化文書の内部構造や構造パターンの文法を知らないユーザでも意図した構造パターンを容易に得ることのみを目的としているからである。したがって、自動作成される構造パターンは、耐久性ある構造パターンであるとは限らない。また、自動作成された構造パターンが耐久性を有するか否かを知るために、ユーザに構造パターンに関する詳しい知識が要求される点に変わりはない。

【0013】更に、構造化文書の内容によっては、将来変更が行われる部分がある程度予測することができる。したがって、構造パターンのうち、将来変更が行われると予測される部分を編集希望項として指定し、当該部分を耐久性ある表現で補完できれば、ユーザは幾通りもある構造パターンから、目的に適った構造パターンのみを迅速に取得することができる。

【0014】この点、ツリー構造を有するUNIX(R)のファイルシステムを扱うユーザ・インターフェース（シェルなど）では、ファイルパス補完機能を提供する。UNIX(R)のファイルパスは各ファイル階層を「/」で区切り、各ディレクトリ、ファイルを文字列として表示する。そして、bash (Bourne Again SHell) を使用する環境下では、例えば、ls /home/userと入力した後、続けてTabキーを押すと、/home/user1、home/user2、…といった具合に、/home/以下のディレクトリやファイルが表示され、ファイルパスが自動的に補完される。しかしながら、このファイルパス補完機能は、ファイルパスを前から後ろへ補完するものであり、ユーザは、パス内の基点となる階層から欲しい情報の存在する階層まで全て辿らなければならない、パス内のある階層を指定し、当該階層のみを自動的に補完することはできない。

【0015】

【特許文献1】

特開平7-225771号公報

【非特許文献1】イレネ ヴァトン (Irene Vatton)、外7名、“アノテーション イン アマヤ (Annotations in Amaya)”、[online]、平成14年12月、W3C、[平成15年1月14日検索]、インターネット<URL:http://www.w3.org/Amaya/User/Annotations.html>

【非特許文献2】 トーマス・エー・フェルプス (Tomas A. Phelps) およびロバート・ウィレンスキー (Robert Wilensky)、“ロバスト イントラ・ドキュメント ロケーションズ (Robust intra-document locations)”、[online]、平成12年 第9回ワールド・ワイド・ウェブ・カンファレンス (the 9th World Wide Web Conference)、[平成15年1月14日検索]、インターネット<URL: <http://www9.org/w9cdrom/312/312.html>>

【0016】

【発明が解決しようとする課題】 よって、幾通りもある複雑な構造パターンが自動的に生成されるシステムの実現が望まれる。特に、生成された構造パターンの中から、最適な構造パターンを容易に選択することができる、構造パターン自動生成システムの実現が望まれる。このようなシステムでは、ユーザには構造パターンに関する詳細な知識が要求されず、また、編集ミスや入力ミスといった誤りを防ぐことができる。また、幾通りもの構造パターンを自動生成し、その中から、ユーザが最適な構造パターンを容易に選択することができれば、構造化文書の様々な変更に対処することが可能となり、ユーザにとって柔軟性の高いシステムが提供される。更に、構造パターンの任意の項を指定し、当該項のみを耐久性のある表現に自動編集する機能の実現が望まれる。このような機能があれば、ユーザは、編集したい項から指したい要素のある項まで全てをたどる手間が省け、目的に適った構造パターンを迅速に取得することができる。

【0017】

【課題を解決するための手段】 本願の発明の概略を説明すれば、以下の通りである。なお、本願発明により生成される構造パターンは、上述の通り、階層指定子および要素パターンとからなる階層指示項を繰り返すことにより、構造化文書内の要素あるいは要素群を指し示す表現である。

【0018】 本願発明のシステムには、編集対象項決定手段、構造パターン候補生成手段、構造パターン候補一覧生成手段が含まれる。また、当該システムは、構造化文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を記憶する記憶装置を含む。本願発明のシステムに、ユーザにより指示された、上記構造化文書内の要素または要素群を指す構造パターンが与えられると、編集

対象項決定手段により、当該構造パターンの項のうち編集対象となる編集対象項が決定される。次に、構造パターン候補生成手段により、構造化文書の文書論理構造情報を基に、編集対象項は異なる表現の項に編集され、構造パターン候補が生成される。そして、構造パターン候補一覧生成手段により、表示指標にしたがって、生成された構造パターン候補が並び替えられ、構造パターン候補一覧が生成される。最後に当該構造パターン候補一覧が表示され、ユーザに提供される。

【0019】また、本発明の他のシステムでは、編集対象である構造パターンの項のうち、編集を希望する編集希望項をユーザから受け付け、当該編集希望項を編集対象項とすることができる。

【0020】また、本発明では、ユーザに選択基準を与えるための表示指標として、構造化文書において、生成された構造パターン候補が指す要素群と指示された構造パターンが指す要素群との包含関係を示す指標、生成された構造パターン候補の特徴を示すカウント可能な特徴量を例示できる。

【0021】また、本発明では、構造化文書の文書論理構造情報として、構造化文書それ自体、または当該構造化文書の文書型定義を例示できる。

【0022】本発明によれば、ユーザは構造化文書内の特定の要素または要素群を指す構造パターンをシステムに対して指示するだけで、幾通りもある複雑な構造パターンの構造パターン候補一覧を得ることができる。したがって、ユーザは入力の手間が省け、入力ミスや編集ミスといった誤りも防ぐことができる。また、構造パターン候補は表示指標にしたがって並び替えられ、構造パターン候補一覧として提供されるため、ユーザは構造パターンに関する特別な知識がなくとも構造パターンを選択する判断基準を得ることができ、目的に合った構造パターンの選択が容易になる。

【0023】また、本発明によれば、ユーザは編集対象とする構造パターンに対して編集を希望する編集希望項を指定することができる。したがって、対象とする構造化文書の内容から将来変更が行われる部分を予測することができる場合には、当該部分を編集希望項として指定することにより、ユーザは幾通りも考え得る構造パターンから、目的に適った構造パターンのみを迅速に取得することができる。

【 0 0 2 4 】

【発明の実施の形態】以下、本発明の実施の形態を図面に基づいて説明する。ただし、本発明は多くの異なる態様で実施することが可能であり、本実施の形態の記載内容に限定して解釈すべきではない。なお、実施の形態の全体を通して同じ要素には同じ番号を付するものとする。

【 0 0 2 5 】以下の実施の形態では、主に方法またはシステムについて説明するが、当業者であれば明らかなとおり、本発明は方法、システムその他、コンピュータ読取可能なプログラムとしても実施できる。したがって、本発明は、ハードウェアとしての実施形態、ソフトウェアとしての実施形態またはソフトウェアとハードウェアとの組合せの実施形態をとることができる。

【 0 0 2 6 】 1. システムのハードウェア構成

図 1 は、本発明の方法を実現するシステムの概要を示した概念図である。本発明のシステムは、スタンドアロンのコンピュータシステムまたは複数のコンピュータシステムで構成されたコンピュータネットワークで実現できる。図 1 (a) はスタンドアロンのコンピュータ構成の概略を示し、図 1 (b) は、コンピュータネットワークの場合を示す。

【 0 0 2 7 】コンピュータシステムには、中央演算処理装置 1 (CPU)、主記憶装置 2 (メインメモリ : RAM)、不揮発性記憶装置 3 (ROM) 等を有し、バス 4 で相互に接続される。バス 4 には、その他コプロセッサ、画像アクセサレータ、キャッシュメモリ、入出力制御装置 (I / O) 等が接続されてもよい。また、バス 4 には、適当なインターフェースを介して外部記憶装置 5、入力装置 6、出力装置 7、通信制御装置 8 等が接続される。その他、一般的にコンピュータシステムに備えられるハードウェア資源を備えることが可能なことは言うまでもない。

【 0 0 2 8 】外部記憶装置 5 は代表的にはハードディスク装置が例示できるが、これに限られず、光磁気記憶装置、光記憶装置、フラッシュメモリ等半導体記憶装置も含まれる。なお、データの読み出しのみに利用し得る CD - ROM 等の読み出し専用記憶装置もデータあるいはプログラム読み出しにのみ適用する場合には外部記憶装置に含まれる。

【0029】入力装置6は、キーボード等の入力装置、マウス等ポインティングデバイスを備えることができる。入力装置6には音声入力装置も含む。出力装置7としては、CRT、液晶表示装置、プラズマ表示装置が例示できる。

【0030】複数のコンピュータシステムで本発明を実現する場合、図1(b)に示すように、各コンピュータシステムは、LAN、WAN等で接続されてもよく、また、インターネットを介して接続されても良い。これら接続に用いられる通信回線は、専用線、公衆回線の何れでも良い。コンピュータシステムには、パーソナルコンピュータ9、ワークステーション10、メインフレームコンピュータ11等各種のコンピュータが含まれる。

【0031】コンピュータシステムが複数接続されたコンピュータネットワークにおいては、一部のプログラムをユーザのコンピュータで、一部のプログラムをリモートコンピュータで分散的に処理を実行できる。またプログラムで利用されるデータは、それがどのコンピュータに記憶されているかは問われない。つまり、データの所在に関する情報（アドレス）が明らかである限り、データあるいはプログラムの格納場所はコンピュータ上の任意の場所とすることができる。各ネットワークコンピュータ間の通信には公知の通信技術を適用でき、たとえばTCP/IP、HTTP等のプロトコルを用いることができる。また、各記憶装置に記録された各ファイル（データあるいはプログラム）の存在箇所（アドレス）は、DNS、URL等を用いて特定できる。なお、本明細書においてインターネットという用語には、イントラネットおよびエクストラネットも含むものとする。インターネットへのアクセスという場合、イントラネットやエクストラネットへのアクセスをも意味する。コンピュータネットワークという用語には、公的にアクセス可能なコンピュータネットワークと私的なアクセスしか許可されないコンピュータネットワークとの両方が含まれるものとする。

【0032】2. 構造パターン候補生成システム

図2は、本実施形態の構造パターン候補生成システムの概略構成を示す図である。図2に示すように、本実施形態の構造パターン候補生成システムは、編集対象項決定手段20、構造パターン候補生成手段21、構造パターン候補一覧生成手段22とを有している。また、本実施形態の構造パターン候補生成システムは、

外部記憶装置 5、入力装置 6、出力装置 7 とを有している。

【0033】外部記憶装置 5 は、ユーザにより指示される構造パターンの基となる構造化文書の文書論理構造情報、およびユーザに選択基準を与えるための表示指標 25 を格納する記憶装置である。本発明では、文書論理構造情報として、構造化文書 23 および当該構造化文書の文書型定義 24 の少なくとも一方を利用する。構造化文書 23 および文書型定義 24 の両方が存在する場合、文書型定義 24 は構造化文書 23 内に記述されてもよい。あるいは、文書型定義 24 を別の文書ファイルとして記録し、構造化文書 23 内でその文書型定義ファイルを参照してもよい。この場合、図 2 では、構造化文書 23 と文書型定義 24 とが同一の記憶装置に格納されているが、必ずしも同一の記憶装置に格納する必要はない。また、表示指標 25 についても、構造化文書 23 または文書型定義 24 と同一の記憶装置に格納する必要はない。

【0034】入力装置 6 は、編集対象とする構造パターンおよび当該構造パターンの編集希望項の指定がなされる装置である。出力装置 7 は、生成される構造パターン候補一覧の表示がなされる装置である。

【0035】入力装置 6 は、上述したように、キーボード等の入力装置、マウス等ポインティングデバイスを備えることができる。編集対象とする構造パターンを指示するために、ユーザは、キーボード等を用いて構造パターンを直接入力してもよい。また、ユーザは、出力装置 7 の画面上に表示される構造化文書等に対し、マウス等を用いて対象とする要素を指示してもよい。この場合、当該要素を指し示す構造パターンは、その固定パス表現が周知の技術により自動生成され、出力装置 7 の画面上に表示される。

【0036】編集対象項決定手段 20 は、前記したハードディスク等外部記憶装置 5 に記憶されたプログラムコードの実行によりその機能が実現され、主に、前記外部記憶装置 5 と前記中央演算処理装置 1 および主記憶装置 2 等のハードウェア資源が利用される。編集対象項決定手段 20 は、上記入力装置 6 を介したユーザの指示によりシステムに与えられた構造パターンについて、構造パターンの項のうち編集対象とする編集対象項を決定する機能を有する。なお、編集対象とすることのできる項の位置および範囲に制限はないため、構造パターンを構成する

項のあらゆる組合せ（１項のみの場合も含む）が編集対象項となる。一方、ユーザにより指定された編集希望項を受付ける場合、編集対象項決定手段２０は、当該編集希望項を編集対象項として決定する。

【００３７】構造パターン候補生成手段２１は、上記ハードディスク等外部記憶装置５に記憶されたプログラムコードの実行によりその機能が実現され、主に、前記外部記憶装置５と前記中央演算処理装置１および主記憶装置２等のハードウェア資源が利用される。構造パターン候補生成手段２１は、編集対象項決定手段２０により決定された各編集対象項について、構造化文書の文書論理構造情報、すなわち、構造化文書２３および当該構造化文書の文書型定義２４の少なくとも一方を基に、当該編集対象項を異なる表現の項に順次編集することにより構造パターン候補を生成する機能を有する。したがって、構造パターン候補生成手段２１により生成される構造パターン候補の数は、編集対象項と置換可能な異なる表現の項の数となる。なお、当該編集方法については後に詳述する。

【００３８】構造化文書２３は、文書の要素が要素名、要素の内容（コンテンツ）、要素の属性、属性名、属性値等で構造化された文書であり、たとえばXML、SGML、HTML等の文書を例示できる。また、文書型定義２４は、構造化文書２３の文書型あるいは文法を定義するデータであり、前記要素の名前、構造、要素内容の形式等を定義し、要素に関する属性を宣言して属性名、属性値の形式、デフォルトの属性値等を定義する。たとえば、XML文書の場合にはDTD、XMLスキーマを例示でき、SGML文書の場合にはDTDを例示できる。

【００３９】構造パターン候補一覧生成手段２２は、前記したハードディスク等外部記憶装置５に記憶されたプログラムコードの実行によりその機能が実現され、主に、前記外部記憶装置５と前記中央演算処理装置１および主記憶装置２等のハードウェア資源が利用される。構造パターン候補一覧生成手段２２は、ユーザに選択基準を与えるための表示指標２５にしたがって、生成された構造パターン候補を並び替え、構造パターン候補一覧を生成する機能を有する。生成された構造パターン候補一覧は、上述したように出力装置７に表示され、ユーザに提供される。

【００４０】本願では、ユーザに選択基準を与える表示指標２５として、構造化

文書 2 3 において、生成された構造パターン候補が指す要素群と指示された構造パターンが指す要素群との包含関係を示す指標、およびまたは生成された構造パターン候補の特徴を示すカウント可能な特徴量を利用する。包含関係を示す指標および特徴を示すカウント可能な特徴量については後に詳述する。

【0 0 4 1】 3. 構造パターン候補生成方法

図 3 は、本発明の構造パターン候補生成システムの全体の動作を示すフローチャートである。ステップ 3 0 において、ユーザにより指示された構造パターンについて、構造パターン候補の項のうち編集対象とする編集対象項を決定する。次に、ステップ 3 1 において、ステップ 3 0 で決定された編集対象項について、構造化文書 2 3 および当該構造化文書の文書型定義 2 4 の少なくとも一方を基に、当該編集対象項を異なる表現の項に編集することにより構造パターン候補を生成する。最後に、ステップ 3 2 において、表示指標 2 5 にしたがって、ステップ 3 1 で生成された構造パターン候補を並び替え、構造パターン候補一覧を生成し、当該構造パターン候補一覧を表示する。以下、各ステップについて、より詳細に説明する。

【0 0 4 2】 図 4 は、編集対象項を決定する方法の一例を説明するフローチャートである。ステップ 4 0 において、ユーザにより指示された構造パターンについて、更に、ユーザにより編集希望項が指定されているか否かを判断する。編集希望項が指定されていない場合、ステップ 4 1 において、編集対象項を求める。本発明は、幾通りもある複雑な構造パターンの自動生成を目的としており、編集対象項の位置および範囲に制限はない。したがって、指示された構造パターンについて、当該構造パターンの項のあらゆる組合せを求め、各々編集対象項として決定する。なお、ステップ 4 1 において、一度に全ての項の組合せを求めてもよく、あるいは、項の組合せが求まる毎に図 3 のステップ 3 1 に進み、その処理の終了後、再度ステップ 4 1 に戻るようにしてもよい。編集希望項が指定されている場合、ステップ 4 2 において、当該編集希望項を編集対象項として決定する。したがって、構造化文書の内容から将来変更される部分が予測できる場合、編集希望項を指定することにより編集対象項を 1 つに特定し、必要な情報だけを迅速かつ的確に取得することが可能となる。

【0043】図5乃至図7は、構造パターン候補を生成する方法の一例を説明するフローチャートである。図5のステップ50において、外部記憶装置5に、ユーザにより指示された構造パターンの基となる構造化文書23が記憶されているか否かを判断する。構造化文書23が記憶されている場合、ステップ51において、構造化文書23を基に、構造パターン候補が生成される。構造化文書23が記憶されていない場合、ステップ52において、外部記憶装置5に上記構造化文書の文書型定義24が記憶されているか否かを判断する。文書型定義24が記憶されている場合、ステップ53において、文書型定義24を基に、構造パターン候補が生成される。外部記憶装置5に構造化文書23および文書型定義24のいずれも記憶されていない場合、ステップ54において、エラーが検出され処理が終了する。ステップ51またはステップ53において構造パターン候補が生成された後、ステップ54において、未処理の編集対象項が残っているか否かを判断する。編集対象項が残っている場合、ステップ50に戻って一連の処理（ステップ50～ステップ54）を繰り返す。編集対象項が残っていない場合、処理は終了する。

【0044】図6は、構造化文書23を用いて構造パターン候補を生成する方法の一例を説明するフローチャートである。構造パターン候補は、編集対象項を異なる表現の項に編集することにより生成される。上述したように、構造パターンの項は、階層指定子と要素パターンからなる。したがって、編集対象項を異なる表現の項に編集することは、この異なる表現の項の階層指定子、および要素パターンを決定することに他ならない。なお、以下の説明では呼称を簡単にするため、異なる表現の項を「補完候補項」と呼ぶこととする。

【0045】まず、ステップ60において、編集対象項決定手段20により決定された編集対象項に基づいて、ユーザにより指示された構造パターンを、編集対象項より前の前項、編集対象項、編集対象項より後の後項の3つに分割する。次に、ステップ61において、ステップ60における分割処理の結果、指示された構造パターンについて前項が存在するか否かを判断する。前項が存在する場合、処理はそのままステップ63へ進む。前項が存在しない場合、ステップ62において、前項をルートに設定する。これは、編集対象項として構造パターンの最上位

の項が選択されたことを意味するためである。ステップ 6 3 において、指示された構造パターンについて、後項が存在するか否か判断する。後項が存在する場合、ステップ 6 5 において、当該後項および構造化文書 2 3 を基に、補完候補項の要素パターンを決定する。後項が存在しない場合、ステップ 6 4 において、構造化文書 2 3 内に存在する全ての要素パターンを抽出し、各々補完候補項の要素パターンとして決定する。なお、ステップ 6 4 または 6 5 において、決定された要素パターンが複数存在する場合、各要素パターンについて以下に説明するステップ 6 6 乃至 7 2 を繰り返し実行する。ステップ 6 4 または 6 5 からステップ 6 6 へ進み、有限数ある階層指示子の 1 つを補完候補項の階層指示子として決定する。なお、全種類の階層指示子について以下に説明するステップ 6 7 乃至 7 1 を繰り返し実行する。ステップ 6 7 において、ステップ 6 6 において決定された階層指示子をステップ 6 4 または 6 5 において決定された要素パターンと連結し、補完候補項を生成する。ステップ 6 8 において、前項、生成された補完候補項、後項を順次連結して構成される構造パターンを、構造化文書 2 3 に対して適用する。ステップ 6 9 において、適用の結果、当該構造パターンが指し示す要素が当該構造化文書 2 3 内に存在するか否か判断する。存在すると判断された場合、ステップ 7 0 において、当該構造パターンを構造パターン候補として登録する。構造パターンが指し示す要素が構造化文書 2 3 内に存在しない場合またはステップ 7 0 から、ステップ 7 1 に進み、全種類の階層指示子についてステップ 6 6 乃至 7 1 の一連の処理が実行されたかどうか判断する。未処理の階層指示子がある場合、ステップ 6 7 へ戻る。全種類の階層指示子について処理が終了している場合、ステップ 7 2 において、更に、ステップ 6 4 または 6 5 において決定された全要素パターンについて、ステップ 6 6 乃至 7 2 の一連の処理が実行されたか否か判断する。未処理の要素パターンが残っている場合、ステップ 6 6 へ戻る。そうでなければ処理は終了する。

【0 0 4 6】図 7 は、文書型定義 2 4 を用いて構造パターン候補を生成する方法の一例を説明するフローチャートである。その内容は、構造化文書 2 3 を用いた構造パターン候補生成の処理の方法とほぼ同様である。但し、文書型定義 2 4 を利用する場合、生成された構造パターンを構造パターン候補として登録するか否

かの判断は、文書型定義 2 4 に照らして当該構造パターンが妥当であるか否か、すなわち、生成された構造パターンが文書型定義 2 4 に定義された構造化文書 2 3 の文書型あるいは文法にそったものであるか否かによって行う（ステップ 8 8）。

【0 0 4 7】 4. 構造パターン候補提供方法の具体例

本発明の構造パターン候補提供システムの具体的な動作を説明する前に、構造化文書 2 3、文書型定義 2 4、および構造パターンを具体例を挙げて以下に説明する。

【0 0 4 8】 図 8（a）に、本発明において利用する構造化文書 2 3 の 1 例として XML 文書 `personnel.xml` を示す。XML 文書の内容は、XML 宣言、文書型定義、XML 文書の本体の 3 つから成り立つ。XML 宣言は、XML のバージョンやその文書で使う文字コードを記述するためのものであり、XML 文書の先頭に記述される。図 8（a）に示す `personnel.xml` の XML 宣言は、XML 勧告のバージョンは 1.0 であり、文書内で使用する文字符号化方式は UTF-8 であることを示している（`personnel.xml` の 1 行目）。文書型定義 2 4 は、上述したように、別の文書ファイルとして記録し、構造化文書 2 3 内でその文書型定義ファイルを参照してもよい。図 8（a）に示す `personnel.xml` は、`personnel.xml` 内で文書型定義 `personnel.dtd`（図 8（b））を参照する（`personnel.xml` の 2 行目）。XML 文書の本体は、タグで区切られた複数の要素から成り立つ（`personnel.xml` の 3 ～ 1 3 行目）。XML 文書では、XML 文書の論理構造を構成する基本単位である要素は、開始タグ、内容、終了タグで構成される。開始タグは、要素の開始を示すマークアップであり、小なり記号「<」に続けて要素型名（タグ名）、大なり記号「>」から構成される。終了タグは、要素の終了を示すマークアップであり、小なり記号とスラッシュ「</」に続けて、要素型名、大なり記号「>」から構成される。内容は、開始タグと終了タグで囲まれた中身となる部分であり、文字データで構成される。内容に、更に、子要素を埋め込むことにより、入れ子の構造、すなわち階層構造を作ることができる。図 8（a）に示される `personnel.xml` は、要素「`personnel`」の内容に子要素「`person`」を埋め込み、更に、その子要素「`person`」の内容に子要素「`name`」を埋め込むことにより 3 階層構造を

作っている。この階層全体はツリー構造をなす。図8(c)にこのツリー構造を図示する。

【0049】構造パターンは、このようなツリー構造をなす構造化文書内の要素を、階層指示子および要素パターンから構成される項を繰り返すことにより指し示す。上述したように、XML文書における構造パターンはXPathである。そこで、XPathを例に、構造パターンの具体的な表現方法を以下に説明する。

【0050】XPathでは、項はロケーションステップと呼ばれ、ロケーションステップは区切り子「/」によって区切られる。すなわち、XPathは、区切り子「/」で区切られた1つ以上のロケーションステップ（以下「ステップ」という）を並べたものであり、このステップは左から右に結合される。そして、各ステップは、コンテキストノード（その時点で処理中のノードのこと）に対して相対的なノード（XML文書を構成するツリーの各要素をいい、厳密には、属性、名前空間名、要素内の文字列等も含む）の集合を順次選択する。例えば、「child::div/child::para」は、コンテキストノードが持つdivという名前の子要素のparaという名前の子要素を選択する。

【0051】ロケーションステップは、基準点、ノードテスト、ゼロ個以上の述語、の3つの部分からなる。その表記は、基準点とノードテストとが二重コロンの区切られ、その後に角括弧で囲まれた式がゼロ個以上続く。基準点は、ロケーションステップを使用して選択するノードとコンテキストノードとの間のツリー関係を指定する。すなわち、基準点はXPathにおける階層指示子である。ノードテストは、ロケーションステップを使用して選択するノードのノード型と展開された名前を指定する。述語は、ロケーションステップを使用して選択するノードの集合を、任意の式を使用して更に細かく選別する。すなわち、ノードテストおよび述語はXPathにおける要素パターンである。例えば、「child::data[2]」は、コンテキストノードの子である要素dataのうち、2番目のdataを選択する。

【0052】XPathでは、階層指示子として13種類の基準点を利用できる。以下、各基準点について説明する。child基準点はコンテキストノードの子ノ

ードを選択する。parent基準点はコンテキストノードの親ノードを選択する。descendant基準点はコンテキストノードの子孫ノードを選択する。子孫ノードとは、子ノードまたは子ノードの子ノード等である。ancestor基準点は、コンテキストノードの祖先ノードを選択する。祖先ノードとは、親ノードまたは親ノードの親ノード等である。following-sibling基準点は、コンテキストノードの後ろにある全ての兄弟ノードを選択する。preceding-sibling基準点は、コンテキストノードの前にある全ての兄弟ノードを選択する。following基準点は、コンテキストノードと同じドキュメント内にあり、ドキュメント順でコンテキストノード以降にある全てのノードを選択する。但し子孫ノードは除く。preceding基準点は、コンテキストノードと同じドキュメント内にあり、ドキュメント順でコンテキストノードより前にある全てのノードを選択する。但し祖先ノードは除く。self基準点は、コンテキストノード自身のみを選択する。descendant-or-self基準点は、コンテキストノードとそのコンテキストノードの子孫ノードを選択する。ancestor-or-self基準点は、コンテキストノードとそのコンテキストノードの祖先ノードを選択する。attribute基準点はコンテキストノードの属性を選択する。namespace基準点は、コンテキストノードの名前空間を選択する。

【0053】図8を用いて、本発明の構造パターン候補提供システムの具体的な動作を説明する。外部記憶装置5には構造化文書23として図8(a)に示すpersonnel.xmlが格納されているものとする。編集対象として、図8(c)の要素name100を指し示す構造パターンがユーザにより指示された場合を考える。上記XPathによれば、要素name100を指し示す固定パスは、/child::personnel[1]/child::person[1]/child::name[1]である。

【0054】第1実施例として、編集希望項として、ユーザにより構造パターンの途中の1項が指定される場合を説明する。ユーザにより指示された構造パターン/child::personnel[1]/child::person[1]/child::name[1]に対し、更に、/child::person[1]がユーザにより編集希望項として指定される。編集対象項決定手段20は/child::person[1]を編集対象項として決定する(ステップ40、42)。

【0055】次に、構造パターン候補生成手段21は、外部記憶装置5に構造化

文書personnel.xmlが格納されていることを確認する（ステップ50）。確認ができると、構造パターン候補生成手段21は、編集対象項決定手段20から受け取った編集対象項情報を基に、構造パターン、/child::personnel[1]/child::person[1]/child::name[1]を分割し、前項に/child::personnel[1]を、後項に/child::name[1]を設定する（ステップ60）。次に、後項が存在することから、構造パターン候補生成手段21は、後項/child::name[1]およびpersonnel.xmlを基に、補完候補項の要素パターン、すなわちタグ名を決定する（ステップ63、65）。図8（c）に示すpersonnel.xmlのツリー構造からも明らかなように、要素name[1]を子にもつ親要素はpersonのみである。そこで構造パターン候補生成手段21は、補完候補項のタグ名をpersonに決定する。次に、決定されたタグ名personと全種類の階層指示子、すなわち13種類の基準点を組合せ、補完候補項を求める（ステップ66、67）。例えば、/descendant::person、/ancestor::person、…という具合である。そして、先程の前項、求められた補完候補項、後項を順次連結し得られる構造パターンを、構造化文書personnel.xmlに適用する（ステップ68）。適用の結果、構造パターンが指し示す要素が構造化文書personnel.xml内に存在すれば、当該構造パターンを構造パターン候補として登録する（ステップ70）。タグ名personを基準点descendantと組み合わせた補完候補項/ancestor::personを例にとると、得られる構造パターンは/child::personnel[1]/descendant::person/child::name[1]である。この構造パターンは、図8（c）の要素name100～102を指し示す。したがって、構造パターン/child::personnel[1]/descendant::person/child::name[1]は、構造パターン候補として登録される。一方、タグ名personを基準点ancestorと組み合わせた補完候補項/ancestor::personを例にとると、得られる構造パターンは/child::personnel[1]/ancestor::person/child::name[1]である。この構造パターンが指し示す要素は、構造化文書personnel.xmlに存在しない。したがって、構造パターン/child::personnel[1]/ancestor::person/child::name[1]は、構造パターン候補として登録されない。このように13種類の全ての基準点についてその組合せを検討すると、最終的に以下の3つの構造パターンが、構造パターン候補として登録される。


```
/child::personnel [1] /descendant::person/child::name [1]
```

```
/child::personnel [1] /descendant-or-self::person/child::name [1]
```

```
/child::personnel [1] /child::person/child::name [1]
```

これら構造パターンは、何れも要素name 1 0 0～1 0 2からなる要素群を指し示す。

【0 0 5 6】第2実施例として、編集希望項として、ユーザにより構造パターンの連続する2項が指示される場合を説明する。2項が1項として表現される、いわゆる縮退の場合である。編集対象として指示された構造パターン/child::personnel [1] /child::person [1] /child::name [1] に対し、更に、ユーザにより、編集希望項として/child::person [1] /child::name [1] が指示される。編集対象項決定手段2 0は/child::person [1] /child::name [1] を編集対象項として決定する（ステップ4 0、4 2）。

【0 0 5 7】次に、構造パターン候補生成手段2 1は、外部記憶装置5に構造化文書personnel.xmlが格納されていることを確認する（ステップ5 0）。確認ができると、構造パターン候補生成手段2 1は、編集対象項決定手段2 0から受け取った編集対象項情報を基に、構造パターン/child::personnel [1] /child::person [1] /child::name [1] を分割し、前項に/child::personnel [1] を設定する（ステップ6 0）。今度は、後項が存在しないことから、構造パターン候補生成手段2 1は、personnel.xmlから起こりえる全ての要素パターン、すなわちタグ名を抽出し、補完候補項のタグ名として決定する（ステップ6 3、6 4）。図8（c）から明らかなように、抽出されるタグ名はpersonnel、person、nameの3つである。タグ名が決定された後の処理、すなわち、各タグ名に対して構造パターン候補生成手段2 1が行う処理は、第1実施例に対して上述したタグ名決定以後の処理と同様である。最終的に以下の構造パターンが構造パターン候補として登録される。

```
/child::personnel [1] /child::person
```

```
/child::personnel [1] /descendant::person
```

```
/child::personnel [1] /descendant-or-self::person
```

```
/child::personnel [1] /descendant::name
```

```

/child::personnel [1] /descendant-or-self::name
/child::personnel [1] /descendant-or-self::personnel
/child::personnel [1] /self::personnel
/child::personnel [1] /ancestor::personnel

```

なお、上記構造パターンのうち、/child::personnel [1] /descendant::nameおよび/child::personnel [1] /descendant-or-self::name以外の構造パターン候補は、元の構造パターンが指す要素name 1 0 0を指さない。このように元の要素を指さない構造パターン候補をユーザに提供するか否かについては、以下に説明する表示指標 2 5 の内容による。

【0 0 5 8】上記 2 つの実施例は、いずれも編集希望項がユーザにより指定される場合であった。編集希望項がユーザにより指定されない場合、プログラムにより擬似的に編集希望項を選択する。すなわち、編集対象項決定手段 2 0 は、指示された構造パターンの項のあらゆる組合せを、各々、編集対象項として決定する。具体例を挙げて説明すると、指示された構造パターンが/child::personnel [1] /child::person [1] /child::name [1] である場合、編集対象項決定手段 2 0 は/child::personnel [1]、/child::person [1]、/child::name [1]、/child::personnel [1] /child::person [1]、/child::person [1] /child::name [1] の 5 つの組合せを求め、各々編集対象項として決定する。

【0 0 5 9】 5. 構造パターン候補一覧生成および表示方法

図 9 は、構造パターン候補一覧を生成する方法の一例を説明するフローチャートである。構造パターン候補一覧生成手段 2 2 は、ステップ 1 1 0 において、外部記憶装置 5 に包含関係を示す表示指標が記憶されているか否か判断する。包含関係を示す表示指標が記憶されている場合、ステップ 1 1 1 において、当該表示指標にしたがって、構造パターン候補生成手段 2 1 によって登録された構造パターン候補を並び替える。なお、表示指標に示される包含関係の何れにも属さない構造パターン候補は、構造パターン候補一覧生成の対象から外される。次に、ステップ 1 1 2 において、特徴量に関する表示指標が記憶されているか否かを判断する。特徴量に関する表示指標が記憶されている場合、ステップ 1 1 3 において、当該特徴量にしたがって構造パターン候補を並び替える。なお、ステップ 1 1 1

で並び替えられた構造パターン候補については、表示指標に示される包含関係ごと、更に並び替えを行う。ステップ 1 1 4 において、並び替えを実行された構造パターン候補を基に、構造パターン候補一覧を生成する。なお、いずれの表示指標も存在しない場合、構造パターン候補一覧生成手段 2 2 は、登録された構造パターン候補全てを単純列挙することにより、構造パターン一覧を作成する。

【0 0 6 0】図 1 0 乃至図 1 3 に、構造パターン候補を表示するユーザ・インターフェースの例を示す。外部記憶装置 5 には、構造化文書 2 3 としてある HTML 文書が格納されているとする。図 1 0 乃至図 1 3 に示す現在の X P a t h 欄 1 2 0 は、編集対象としてユーザに指示された構造パターンを表示する。また、構造パターンの一部の項の反転表示は、当該項がユーザによって指示された編集希望項であることを示す。更に、図中に示されるチェックボックス 1 2 1 ~ 1 2 4 は、自動生成される構造パターン候補のうち、表示することを希望する構造パターン候補を指示する際に利用されるものである。

【0 0 6 1】図 1 0 は、構造パターン候補が指す要素群と元の構造パターン（現在の X P a t h）が指す要素群との包含関係を示す表示指標 2 5 にしたがって生成された構造パターン候補一覧を表示する画面の 1 例を示す。当該表示指標 2 5 にしたがって、構造パターン候補生成手段 2 1 によって登録された構造パターン候補は 4 つに並び替えられる。現在選択されている要素を含む欄 1 2 5 には、構造パターン候補の指す要素群が元の構造パターンが指す要素群を含むような構造パターン候補が表示される。なお、表示領域の制限から、構造パターンの全体でなく、該当する構造パターン候補の補完候補項部分のみを表示してもよい。図 1 0 には 2 つの補完候補項、/descendant::TD および /descendant-or-self::TD が表示されている。現在選択されている要素と同じ要素を指す欄 1 2 6 には、構造パターン候補の指す要素群が、元の構造パターンが指す要素群と同じになるような構造パターン候補が表示される。本実施例では該当する構造パターン候補は存在しない。現在選択されている要素を一部含む欄 1 2 7 には、構造パターン候補の指す要素群が、元の構造パターンが指す要素群と一部重なるような構造パターン候補が表示される。本実施例では該当する構造パターン候補は存在しない。現在選択されている要素を含まない欄 1 2 8 には、構造パターン候補の指す要素群が

、元の構造パターンが指す要素群を全く含まないような構造パターン候補が表示される。図 1 0 に示すように当該欄に分類される構造パターン候補は 6 つ存在する。

【0 0 6 2】図 1 1 は、上述した図 1 0 に示すユーザ・インターフェースの下、現在の X P a t h として異なる X P a t h を指示した例を示す。図 1 1 に示すように、構造パターン候補生成手段 2 1 によって登録された 6 つの構造パターン候補は、現在選択されている要素と同じ要素を指す欄 1 2 6 および現在選択されている要素を一部含む欄 1 2 7 に分類される。

【0 0 6 3】図 1 0 および図 1 1 に示す実施例によれば、自動生成された複数の構造パターン候補は、包含関係を示す表示指標にしたがって並び替えられ、ユーザに提供される。ユーザは、構造パターン候補と現在の X P a t h との関係を、自ら解析することなく知ることができる。したがって、当該並び替えを基に、幾通りもある構造パターン候補の中から、対象とする構造化文書の変更特性等に最も適した構造パターンを、ユーザは容易に選択することができる。

【0 0 6 4】図 1 2 は、更に、特徴量に関する表示指標にしたがって生成された構造パターン候補一覧を表示する画面の 1 例を示す。ここで、特徴量とは、構造パターン候補の特徴を示すカウント可能な量を意味する。図 1 2 に示すユーザ・インターフェースでは、ロケーションステップの数、すなわち、項の数を特徴量として構造パターン候補を並び替える。項の数は情報量を示し、項の数が多いと厳密な X P a t h、項の数が少ないと曖昧な X P a t h といえる。それゆえ、項の数は構造パターン候補の耐久性の強度を示す指標ともなる。ここではロケーションステップの数を特徴量として前述の表示指標に加え、ユーザの構造パターン候補の選択を助ける。なお、本実施例ではユーザによる編集希望項の指示はない。また、包含関係に関する表示指標は、現在の X P a t h が指す要素群と同じ要素群を指す構造パターン候補のみを表示するよう指示する。当該指示は、好ましくは、上述したチェックボックス 1 2 2 の選択によりなされる。このように、ユーザは必要な情報のみを表示するようになれる。

【0 0 6 5】図 1 3 は、ノード（要素）の数を特徴量とする表示指標にしたがって生成された構造パターン候補一覧を表示する画面の 1 例を示す。本実施例にお

いても、ユーザによる編集希望項の指示はない。ただし、包含関係に関する表示指標は、現在のX P a t hが指す要素群を含む要素群を指す構造パターン候補のみを表示するよう指示する。したがって、表示される構造パターン候補が指す要素群は、いずれも現在のX P a t hが指す要素群を含んでいる。しかしながら、表示される構造パターン候補が指す要素の数は各々異なり、指し示す要素数が多いほど、現在のX P a t hに対して曖昧なX P a t hであるといえる。それゆえ、構造パターン候補が指す要素の数は、構造パターン候補の耐久性の強度を示す指標ともなる。ここではノードの数を特徴量として前述の表示指標に加え、ユーザの構造パターン候補の選択を助ける。

【0066】図12および図13に示す実施例では、包含関係を示す表示指標と特徴量に関する表示指標を同時に利用したが、特徴量に関する表示指標のみを利用できることはいうまでもない。また、図12および図13に示すように、項の数／ノードの数を構造パターン候補と共に表示することにより、よりユーザ・フレンドリーな構造パターン候補提供システムを提供することができる。

【0067】図14は、編集希望項をユーザが選択する際の支援機能を示す図である。原則として、編集を希望する項の位置と項の数は任意でよい。しかしながら、図14に示すように、表示指標が現在のX P a t hが指す要素群と同じ要素群を示す構造パターン候補のみを表示するよう指示する場合、選択できる編集希望項は制限される。そこで、本実施例では、表示指標の指示にしたがって選択可能な編集希望項を長方形の枠で囲み、ユーザの選択を支援する。

【0068】以上のように、本発明者によってなされた発明を発明の実施の形態に基づき具体的に説明したが、本発明は前記実施の形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変更可能である。

【0069】また、前記実施の形態では、主にシステムおよび方法について説明したが、これらシステムの機能、方法はコンピュータに可読なプログラムにより実現されるため、このようなコンピュータ可読なプログラムが記録された媒体が前記実施の形態において開示されていることは明らかである。なお、媒体には、たとえばハードディスク等記憶装置に内蔵された媒体のみならず、CD-ROM、DVD-ROM、フラッシュメモリ、フレキシブル・ディスク等、記憶装置と

は別個に取り引きされるものを含むことは勿論である。

【0070】まとめとして、本発明の構成に関して以下の事項を開示する。

(1) 構造化文書内の要素または要素群を指す構造パターンの候補を生成する方法であって、構造化文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を用意するステップと、ユーザにより指示された、構造化文書内の要素または要素群を指す構造パターンを受付けるステップと、構造パターンを構成する項のうち編集対象となる編集対象項を決定するステップと、文書論理構造情報を基に、編集対象項を異なる表現の項に編集することにより構造パターン候補を生成するステップと、表示指標にしたがって、生成された構造パターン候補を並び替え、構造パターン候補一覧を生成するステップと、を含む構造パターン候補を生成する方法。

(2) 前記編集対象項を決定するステップは、構造パターンについてユーザにより指示された編集希望項を受付けるステップと、当該編集希望項を編集対象項として決定するステップとを含む、(1)記載の構造パターン候補を生成する方法。

(3) 前記表示指標は、構造化文書において、生成された構造パターン候補が指す要素群とユーザにより指示された構造パターンが指す要素群との包含関係を示す指標を含む、(1)または(2)いずれか記載の構造パターン候補を生成する方法。

(4) 前記表示指標は、生成された構造パターン候補の特徴を示すカウント可能な特徴量を含む、(1)乃至(3)いずれか記載の構造パターン候補を生成する方法。

(5) 前記文書論理構造情報を用意するステップは、構造化文書および当該構造化文書の文書型定義の少なくとも一方を用意するステップを含む、(1)乃至(4)いずれか記載の構造パターン候補を生成する方法。

(6) 構造化文書内の要素または要素群を指す構造パターンの候補を生成するシステムであって、構造化文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を記憶する記憶装置と、ユーザにより指示された、構造化文書内の要素または要素群を指す構造パターンを受付ける手段と、構造パターンを構

成する項のうち編集対象となる編集対象項を決定する手段と、文書論理構造情報を基に、編集対象項を異なる表現の項に編集することにより構造パターン候補を生成する手段と、表示指標にしたがって、生成された構造パターン候補を並べ替え、構造パターン候補一覧を生成する手段と、を含む構造パターン候補を生成するシステム。

(7) 編集対象項を決定する手段は、構造パターンについてユーザにより指示された編集希望項を受付ける手段と、当該編集希望項を編集対象項として決定する手段とを含む、(6) 記載の構造パターン候補を生成するシステム。

(8) 表示指標は、構造化文書において、生成された構造パターン候補が指す要素群とユーザにより指示された構造パターンが指す要素群との包含関係を示す指標を含む、(6) または (7) いずれか記載の構造パターン候補を生成するシステム。

(9) 表示指標は、生成された構造パターン候補の特徴を示すカウント可能な特徴量を含む、(6) 乃至 (8) いずれか記載の構造パターン候補を生成するシステム。

(10) 記憶装置は、文書論理構造情報として、構造化文書および当該構造化文書の文書型定義の少なくとも一方を記憶する、(6) 乃至 (9) いずれかの請求項記載の構造パターン候補を生成するシステム。

(11) 構造化文書内の要素または要素群を指す構造パターンの候補を生成するためのプログラムであって、該プログラムがコンピュータに、構造化文書の文書論理構造情報およびユーザに選択基準を与えるための表示指標を用意する機能、ユーザにより指示された、構造化文書内の要素または要素群を指す構造パターンを受付ける機能、構造パターンを構成する項のうち編集対象となる編集対象項を決定する機能、文書論理構造情報を基に、編集対象項を異なる表現の項に編集することにより構造パターン候補を生成する機能、および表示指標にしたがって、生成された前記構造パターン候補を並べ替え、構造パターン候補一覧を生成する機能を実現させる構造パターン候補生成プログラム。

(12) XML文書内の要素または要素群を指すXPathの候補を生成する方法であって、XML文書の文書論理構造情報およびユーザに選択基準を与えるた

めの表示指標を用意するステップと、ユーザにより指示された、XML文書内の要素または要素群を指すX P a t hを受付けるステップと、X P a t hを構成するロケーションステップのうち編集対象となるロケーションステップを決定するステップと、文書論理構造情報を基に、決定されたロケーションステップを異なる表現のロケーションステップに編集することによりX P a t h候補を生成するステップと、表示指標にしたがって、生成されたX P a t h候補を並び替え、X P a t h候補一覧を生成するステップと、を含むX P a t h候補を生成する方法。

【図面の簡単な説明】

【図 1】 本発明の方法を実現するシステムの概要を示した概念図であり、（a）はスタンドアロンのコンピュータの概略を示し、（b）は、コンピュータネットワークの構成の概略を示す。

【図 2】 本発明の構造パターン候補生成システムの一例を示したブロック図である。

【図 3】 本発明の構造パターン候補を生成する方法の一例を示す全体フローチャートである。

【図 4】 編集対象項を決定する方法の一例を示すフローチャートである。

【図 5】 構造パターン候補を生成する方法の一例を示すフローチャートである。

【図 6】 構造化文書により構造パターン候補を生成する方法の一例を示すフローチャートである。

【図 7】 文書型定義により構造パターン候補を生成する方法の一例を示すフローチャートである。

【図 8】 （a）は、構造化文書の一例を示す図である。（b）は、（a）の構造化文書の文書型定義を示す図である。（c）は、（a）の構造化文書のツリー構造を示す図である。

【図 9】 構造パターン候補を表示する方法の一例を示すフローチャートである。

【図 1 0】 包含関係を示す表示指標にしたがって生成された構造パターン候補一覧を表示する画面の一例を示した図である。

【図 1 1】 包含関係を示す表示指標にしたがって生成された構造パターン候補一

覧を表示する画面の一例を示した図である。

【図 1 2】ロケーションステップ数を特徴量とする表示指標にしたがって生成された構造パターン候補一覧を表示する画面の一例を示した図である。

【図 1 3】ノード数を特徴量とする表示指標にしたがって生成された構造パターン候補一覧を表示する画面の一例を示した図である。

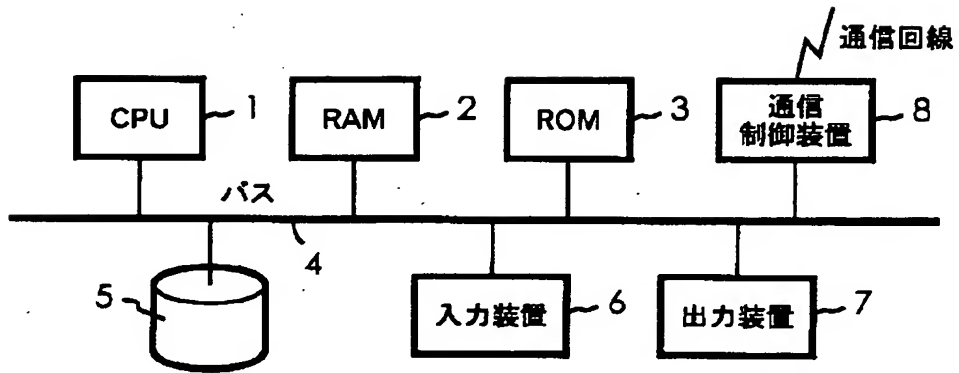
【図 1 4】編集希望項選択を支援する画面の一例を示した図である。

【図 1 5】（a）は、ある構造化文書のツリー構造を示した図である。（b）および（c）は、各々、（a）のある構造化文書が変更された後の、当該構造化文書のツリー構造を示した図である。

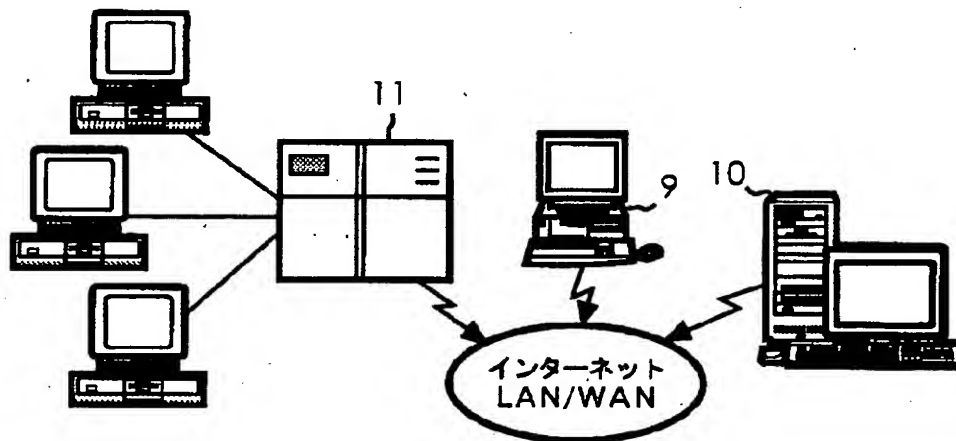
【書類名】 図面

【図 1】

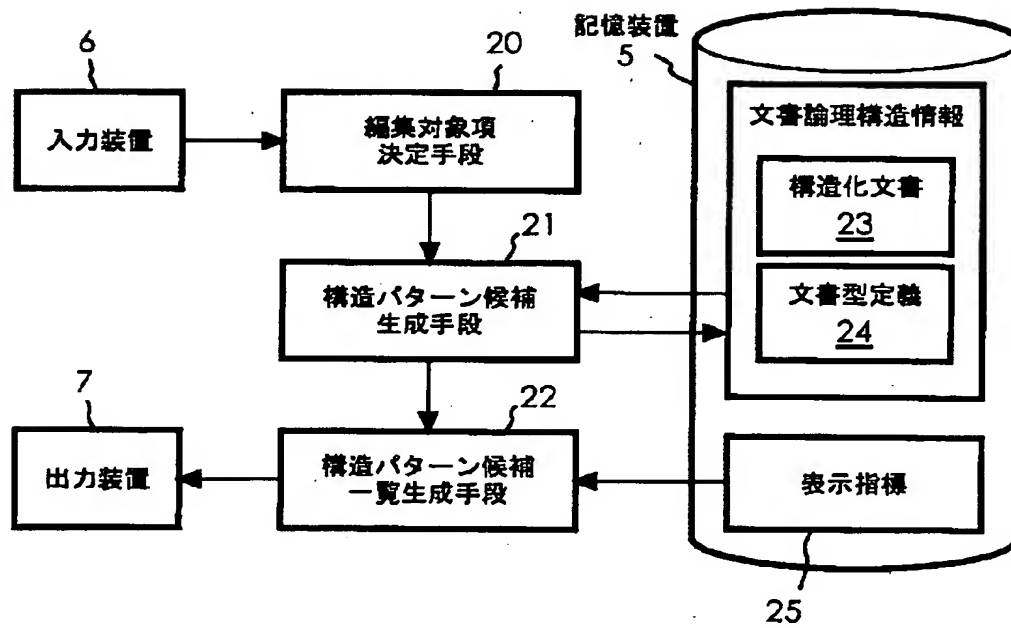
(a)



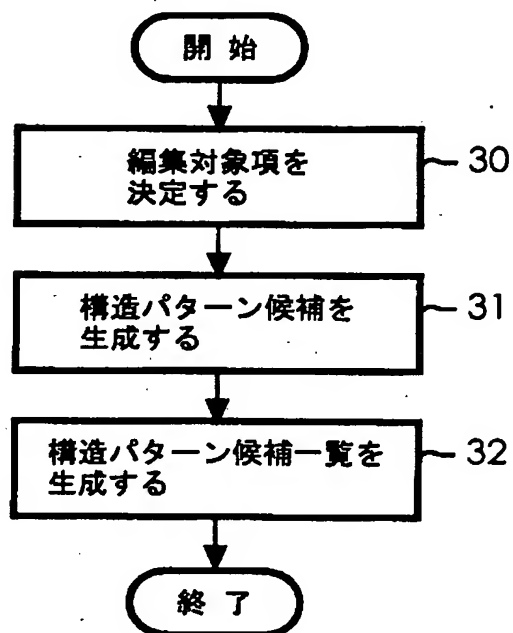
(b)



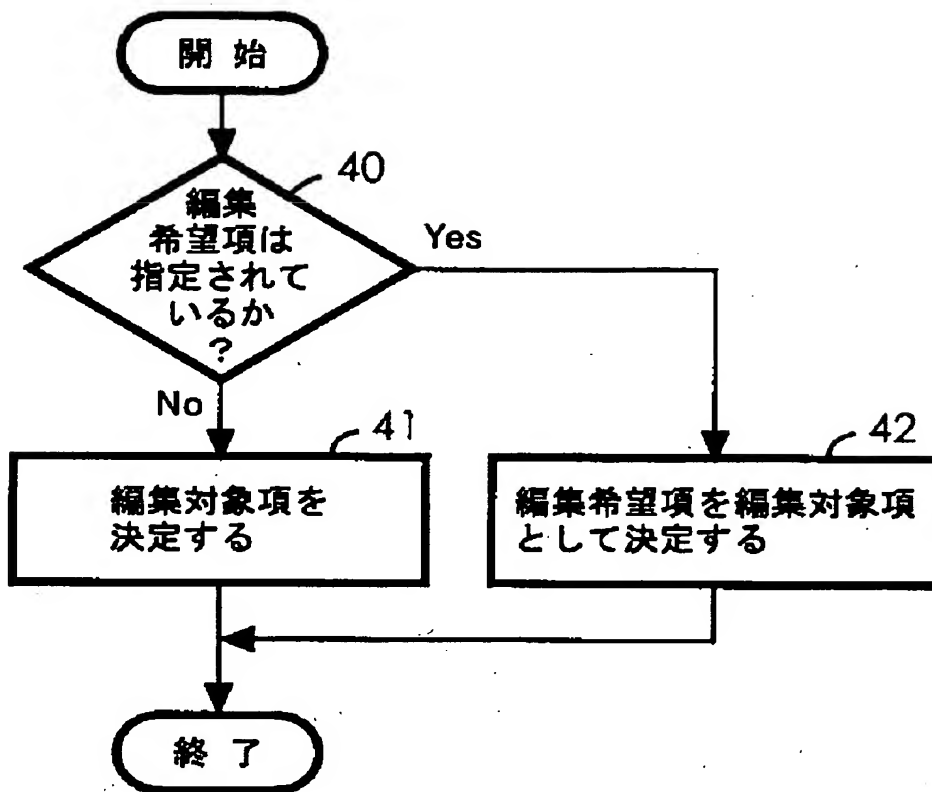
【図 2】



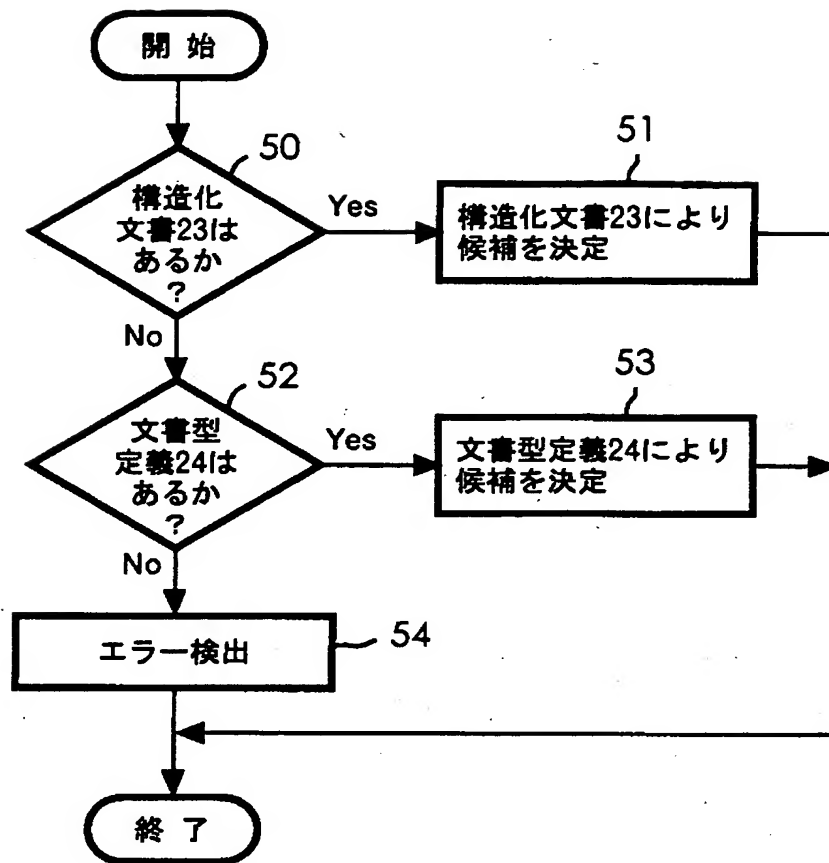
【図 3】



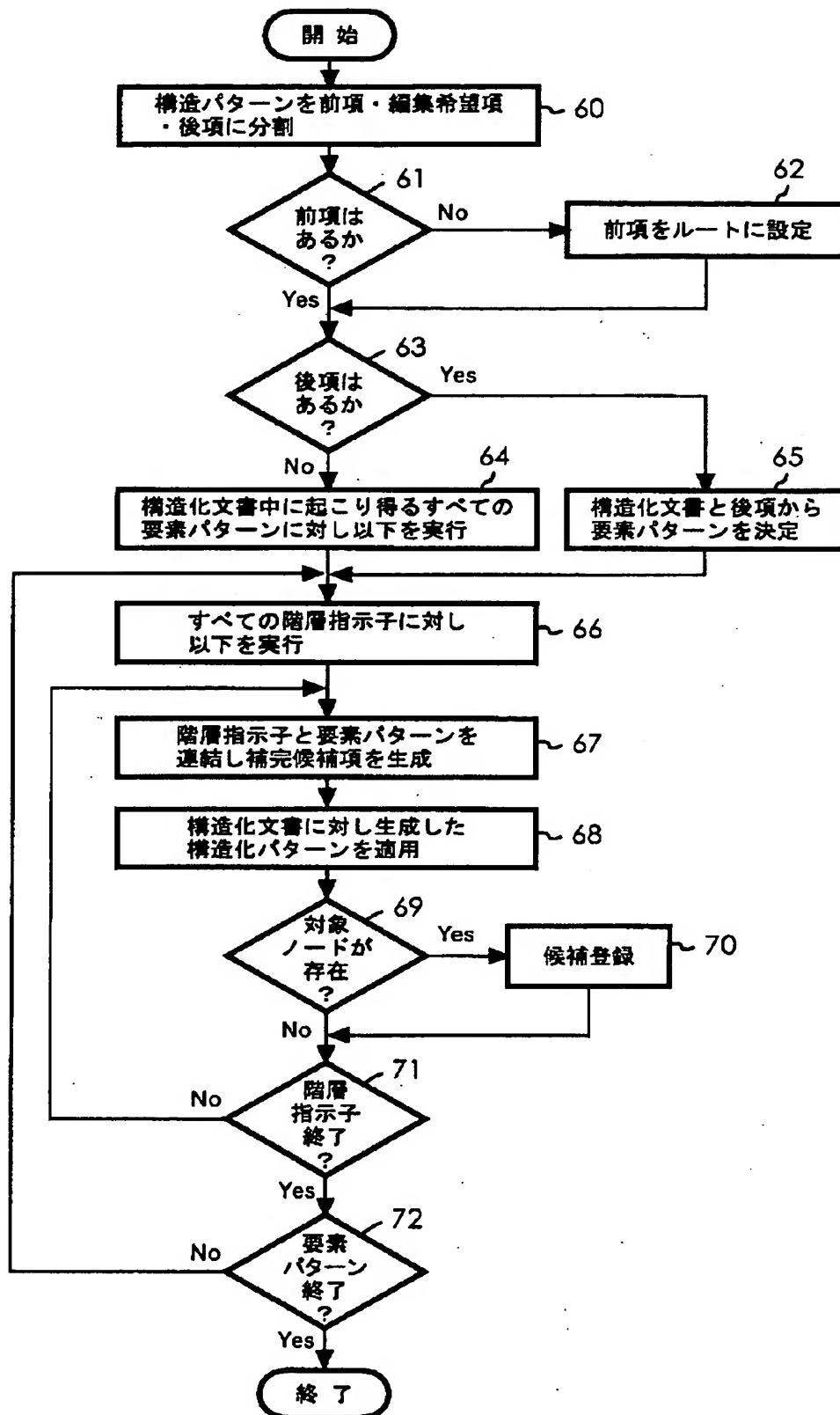
【図 4】



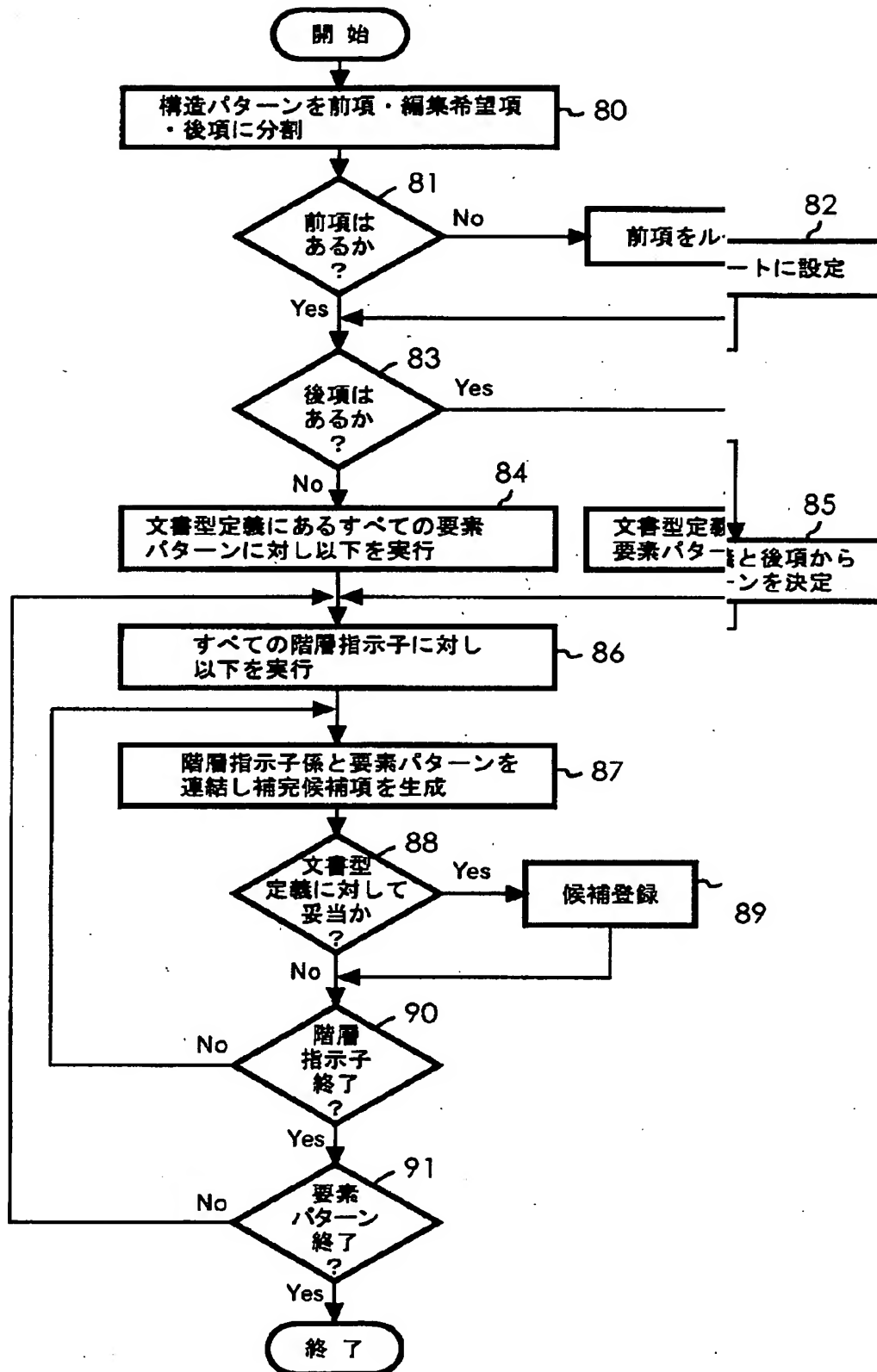
【図 5】



【図 6】



【図 7】



【図 8】

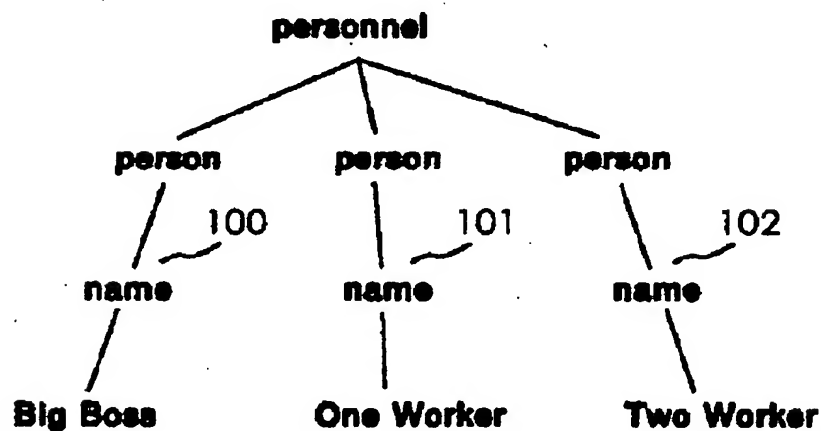
(a) personnel.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE personnel SYSTEM "personnel.dtd">
<personnel>
  <person>
    <name>Big Boss</name>
  </person>
  <person>
    <name>One Worker</name>
  </person>
  <person>
    <name>Two Worker</name>
  </person>
</personnel>
```

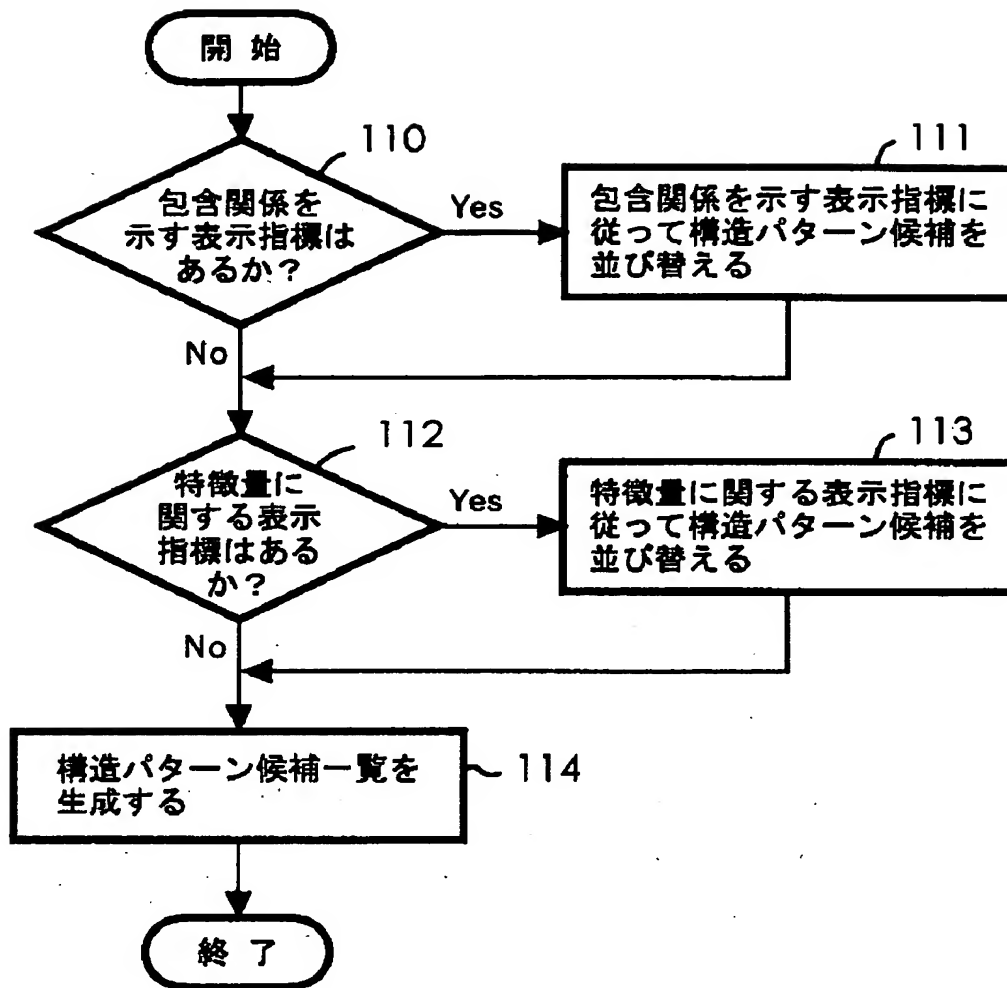
(b) personnel.dtd

```
<!ELEMENT personnel(person+)>
<!ELEMENT person(name+)>
<!ELEMENT name(#PCDATA)>
```

(c)



【図9】



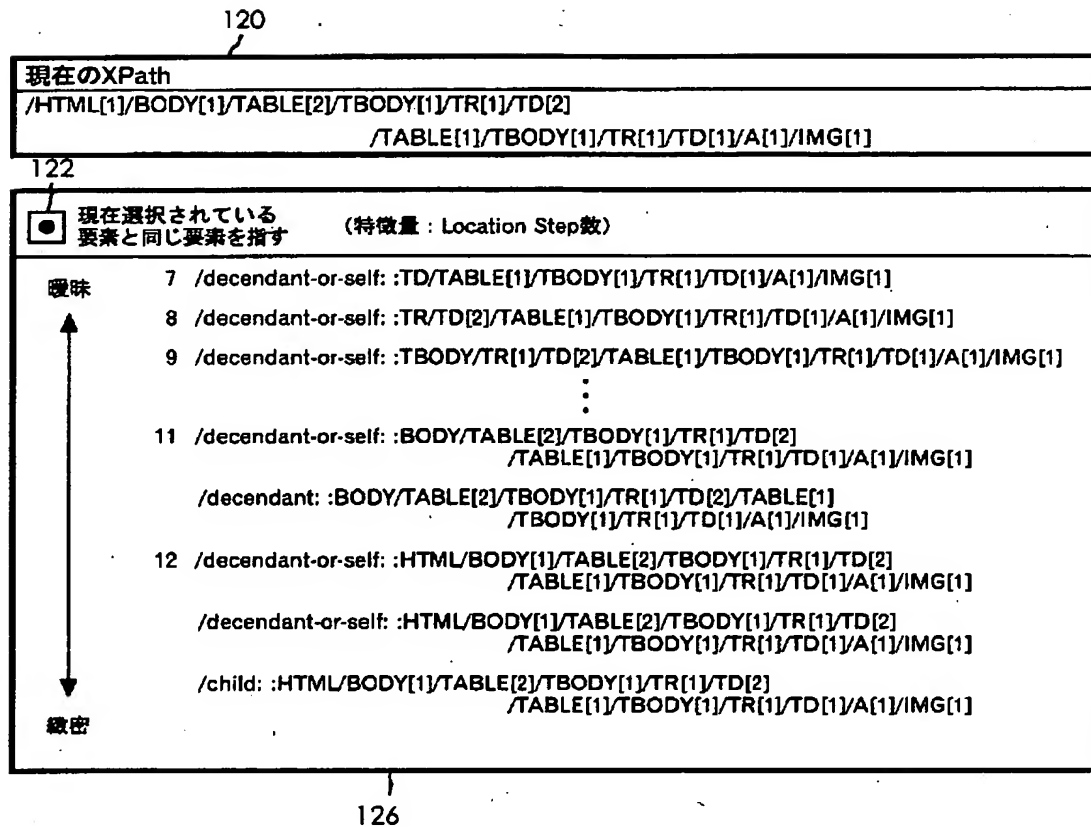
【図 10】

<div>120</div> <div>現在のXPath</div> <div>/HTML[1]/BODY[1]/TABLE[2]/TBODY[1]/TR[1]/TD[2]</div> <div>/TABLE[1]/TBODY[1]/TR[3]/TD[1]/TABLE[1]/TBODY[1]/TR[3]/TD[3]/A[1]</div>			
121	122	123	124
<input checked="" type="checkbox"/> 現在選択されている要素を含む	<input type="checkbox"/> 現在選択されている要素と同じ要素を指す	<input checked="" type="checkbox"/> 現在選択されている要素を一部含む	<input type="checkbox"/> 現在選択されている要素を含まない
descendant: :TD descendant-or-self: :TD			self: :TD following: :TD descendant: :SPAN descendant-or-self: :SPAN ancestor-or-self: :TD preceding: :TD
125	126	127	128

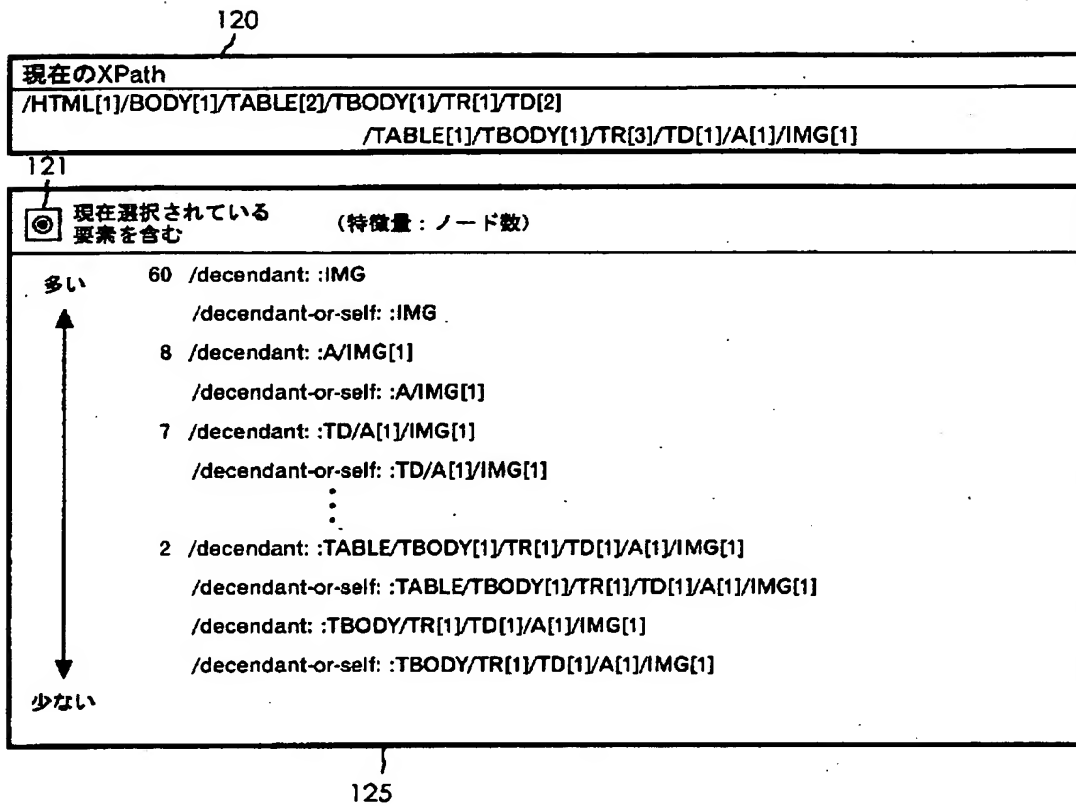
【図 11】

<div>120</div> <div>現在のXPath</div> <div>/HTML[1]/BODY[1]/TABLE[2]/TBODY[1]/TR[1]/TD[2]/TABLE[1]</div> <div>/TBODY[1]/TR[3]/TD[1]/TABLE[1]/TBODY[1]/TR/TD/following: :TR</div>			
121	122	123	124
<input checked="" type="checkbox"/> 現在選択されている要素を含む	<input type="checkbox"/> 現在選択されている要素と同じ要素を指す	<input checked="" type="checkbox"/> 現在選択されている要素を一部含む	<input type="checkbox"/> 現在選択されている要素を含まない
	ancestor-or-self: :TR ancestor: :TR preceding: :TR	following: :FORM following: :TR parent: :TR	
125	126	127	128

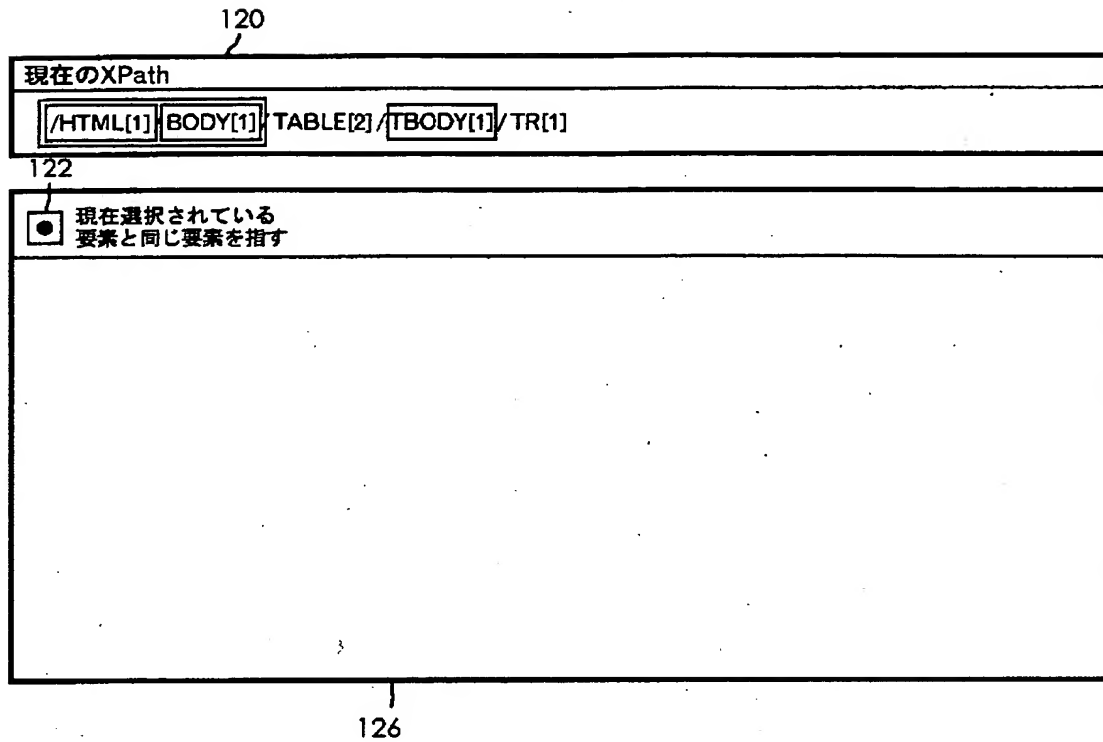
【図 12】



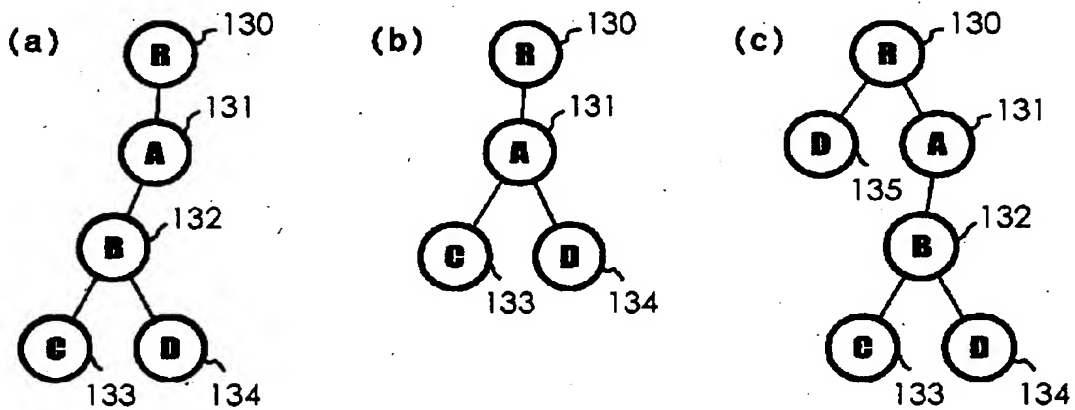
【図 13】



【図 1 4】



【図 1 5】



【書類名】 要約書

【要約】

【課題】幾通りもある複雑な構造パターンの候補を自動生成し、目的に合った構造パターンの選択を容易にする構造パターン候補生成システムを提供する。

【解決手段】ユーザにより指示された、構造化文書の文書論理構造中の要素または要素群を指す構造パターンがシステムに与えられると、編集対象階項決定手段は、当該構造パターンを構成する項のうち、編集対象となる編集対象項を決定する。構造パターン候補生成手段は、構造化文書の文書論理構造情報を基に、編集対象項を異なる表現の項に編集し、構造パターン候補を生成する。構造パターン候補一覧生成手段は、ユーザに選択基準を与えるための表示指標に従って、生成された構造パターン候補を並び替え、構造パターン候補一覧を生成する。構造パターンについてユーザにより指定された編集希望項がシステムに与えられる場合、編集対象項決定手段は、当該編集希望項を編集対象項として決定する。

【選択図】 図 2

認定・付加情報

特許出願の番号	特願2003-022380
受付番号	50300149473
書類名	特許願
担当官	第七担当上席 0096
作成日	平成15年 1月31日

<認定情報・付加情報>

【提出日】	平成15年 1月30日
-------	-------------

出 願 人 履 歴 情 報

識別番号 [390009531]

1. 変更年月日 2002年 6月 3日

[変更理由] 住所変更

住 所 アメリカ合衆国10504、ニューヨーク州 アーモンク ニ
ュー オーチャード ロード

氏 名 インターナショナル・ビジネス・マシーンズ・コーポレーショ
ン